Nonlinear system identification under various prior knowledge

Zygmunt Hasiewicz, Przemysław Śliwiński, Grzegorz Mzyk

The Institute of Computer Engineering Control and Robotics
Wrocław University of Technology, POLAND

IFAC'08, Seoul, July 9th, 2008

Approaches to system identification

- parametric (traditional)
- nonparametric
- parametric-nonparametric (combined)
- semiparametric

Static nonlinearity

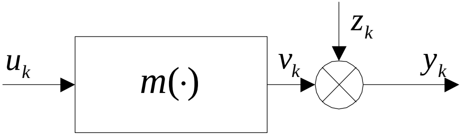


Figure: Static nonlinear element

$$R(u) = E\{y_k | u_k = u\} = E\{m(u_k) + z_k | u_k = u\} = m(u)$$

Two kinds of knowledge

1. parametric, (shape of formula describing $m(u, c^*) = m(u)$)

$$m(u,c) = c_1 f_1(u) + c_2 f_2(u) + \ldots + c_p f_p(u)$$

or

$$m(u,c) = f_1(u,c_1) \circ f_2(u,c_2) \circ \ldots \circ f_p(u,c_p)$$

e.g.

$$m(u,c) = c_1 + c_2 u + c_3 u^2 \qquad \text{or} \qquad m(u,c) = c_1(\sin c_2 u + c_3 e^{c_4 u})$$

2. non-parametric, (measurements)

$$\{(u_k, y_k)\}_{k=1}^{N}$$

The classical approach (parametric)

$$\widehat{c}_N = \arg\min_c \sum_{k=1}^{N} (y_k - m(u_k, c))^2$$

Features:

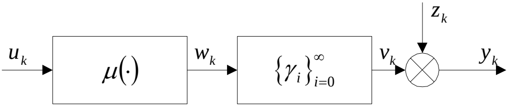
- fast convergence to the optimal model in the class $m(u,c)$ (under rich a priori knowledge)
- but, the risk of systematic approximation error when the model is bad
- complicated and badly conditioned computations (linear or nonlinear least squares)

Nonparametric estimates

Orthogonal expansion, kernel regression

- are based on the measurements only
- does not involve the unknown characteristic must belong to the finite dimmensional class ($p \longrightarrow \infty$ as $N \longrightarrow \infty$)
- converge to the true characteristic
- are computationally simple
- have more deegrees of freedom (the choice of tunning parameters and basis functions)

Hammerstein system



$$R(u) = E\{y_k | u_k = u\} = E\left\{\sum_{i=0}^{\infty} \gamma_i \mu(u_{k-i}) + z_k | u_k = u\right\} = \gamma_0 \mu(u) + \zeta$$

Nonparametric algorithms

Assumptions

- The nonlinear characteristic $m(u)$ can be an *arbitrary* function, square integrable on $[-1,1]$, and *e.g.*:
  - differentiable
  - continuous
  - piecewise-smooth
- There is a set of *sorted* input-output measurements $\{u_l, y_l\}$, $l = 1, \ldots, k$.

Remark

The class of admissible characteristic is now so ample that it **cannot** *be represented by any parametric model*

Orthogonal series basics

Observation

Any nonlinearity $\mu(u)$ has its orthogonal expansion

$$\begin{aligned}\mu(u) &= \alpha_0 + \alpha_1 \cdot p(u) + \cdots + \alpha_K \cdot p_K(u) + \cdots \textit{ ad infinitum} \\ &= \sum_{i=0}^{\infty} \alpha_i \cdot p_i(u)\end{aligned}$$

where $\{p_i\}, i = 0, 1, \ldots$ is an orthogonal basis in $L^2[-1,1]$ and

$$\alpha_i = \langle \mu, p_i \rangle = \int_{-1}^{1} \mu(u) \cdot p_i(u)\, du$$

Orthogonal series estimate

A generic orthogonal series algorithm

An orthogonal series algorithm has a form

$$\hat{\mu}(u) = \sum_{i=0}^{K(k)} \hat{\alpha}_i \cdot p_i(u)$$

where $K(k)$ is an non-decreasing number sequence and

$$\hat{\alpha}_i = \sum_{l=1}^{k} \int_{u_{l-1}}^{u_l} y_l \cdot p_i(u)\, du$$

MISE error

The performance of the algorithm is measured by a mean integrated square error

$$\text{MISE}\,\hat{\mu} = E\int_{-1}^{1}\left(\mu\left(u\right)-\hat{\mu}\left(u\right)\right)^2 du$$

Error decomposition

$$\text{MISE}\,\hat{\mu} = \overbrace{\sum_{i=K(k)+1}^{\infty}\alpha_i^2}^{\text{approx}^2\,\mu^K\text{ - deterministic error}} + \overbrace{\overbrace{\sum_{i=0}^{K(k)}\text{bias}^2\hat{\alpha}_i}^{\text{bias}^2\,\hat{\mu}} + \overbrace{\sum_{i=0}^{K(k)}\text{var}\,\hat{\alpha}_i}^{\text{var}\,\hat{\mu}}}^{\text{stochastic errors}}$$

Example I – Legendre polynomial series estimate

Legendre polynomial estimate

Legendre polynomial basis is *recursively* defined as

$$p_i(u) = \sqrt{\frac{2i+1}{2}} \cdot P_i(u)$$

where

$$P_i(u) = \frac{2i-1}{i} \cdot xP_{i-1}(u) - \frac{i-1}{i} \cdot P_{i-2}(u)$$

with

$$P_1(u) = u \text{ and } P_0(u) = 1$$

Example II – Chebyshev polynomial series estimate

Chebyshev polynomial estimate

Chebyshev polynomial basis is *recursively* defined as

$$p_i(u) = \sqrt{\frac{1}{1-u^2}} \cdot P_i(u)$$

where

$$P_i(u) = 2xP_{i-1}(u) - P_{i-2}(u)$$

with

$$P_1(u) = u \text{ and } P_0(u) = 1$$

Convergence & rates

Convergence

If $K(k) \to \infty$ and $K(k)/k \to 0$ then

$$\mathrm{MISE}\,\hat{\mu} \to 0 \text{ as } k \to \infty.$$

Convergence rate

Let λ be a number of derivatives of μ. If $K(k) = k^{\frac{1}{2\lambda+1}}$ then

$$\mathrm{MISE}\,\hat{\mu} \sim k^{-\frac{2\lambda}{2\lambda+1}}.$$

- the smoother nonlinearity the faster convergence
- the rate can be established for smooth nonlinearities only

Example III - Wavelet series estimate

Any nonlinearity can be represented in a multiresolution form

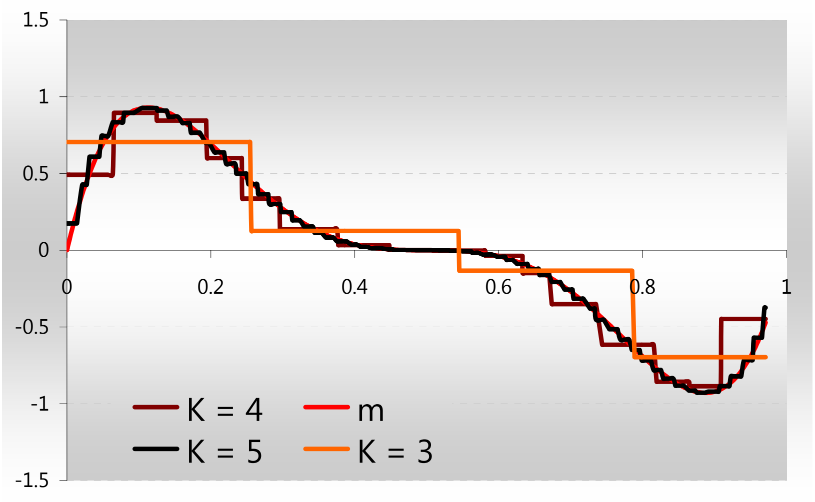
$$\mu(u) = \overbrace{\sum_{n=0}^{2^M-1} \alpha_{Mn} \cdot \varphi_{Mn}(u)}^{\text{A 'crude' approximation}} + \overbrace{\sum_{n=0}^{2^M-1} \beta_{Mn} \cdot \psi_{Mn}(u)}^{\text{details at the resolution } 2^M} + \cdots$$

$$+ \overbrace{\sum_{n=0}^{2^{K-1}-1} \beta_{K-1,n} \cdot \psi_{K-1,n}(u)}^{\text{details at the resolution } 2^{K-1}} + \cdots \textit{ ad infinitum}$$

where

$$\alpha_{Mn} = \int_{-1}^{1} \mu(u) \cdot \varphi_{Mn}(u)\, du \text{ and } \beta_{mn} \int_{-1}^{1} \mu(u) \cdot \psi_{mn}(u)\, du$$

Example III - multiresolution approximation



Example III - Wavelet series estimate

A generic wavelet estimate

The wavelet estimate is of the form

$$\hat{\mu}(u) = \sum_{n=0}^{2^M-1} \hat{\alpha}_{Mn} \cdot \varphi_{Mn}(u) + \sum_{m=M}^{K(k)-1} \sum_{n=0}^{2^m-1} \hat{\beta}_{mn} \cdot \psi_{mn}(u)$$

where

$$\hat{\alpha}_{Mn} = \sum_{l=1}^{k} y_l \cdot \int_{u_{l-1}}^{u_l} \varphi_{Mn}(u)\, du \text{ and } \hat{\beta}_{mn} = \sum_{l=1}^{k} y_l \cdot \int_{u_{l-1}}^{u_l} \psi_{mn}(u)\, du$$

- $\varphi(u)$ and $\psi(u)$ can be from *Haar* or *Cohen-Daubechies-Vial* family...

Convergence & rates

Convergence and its rate

If $K(k) \to \infty$ and $2^{K(k)}/k \to 0$ then

$$\text{MISE}\,\hat{\mu} \to 0 \text{ as } k \to \infty.$$

Let λ be a number of derivatives of μ. If $K(k) = \frac{1}{2\lambda+1}\log_2 k$ then

$$\text{MISE}\,\hat{\mu} \sim k^{-\frac{2\lambda}{2\lambda+1}}$$

Let $\mu(u)$ has a finite number of jumps. If $K(k) = \frac{1}{2}\log_2 k$ then

$$\text{MISE}\,\hat{\mu} \sim k^{-\frac{1}{2}}$$

- the smoother nonlinearity the faster convergence
- the rate can be established also for **discontinuous** nonlinearities!

Kernel estimate

A generic kernel estimate

The algorithm based on kernels has the generic form

$$\hat{\mu}(u) = \frac{\sum_{l=1}^{k} y_l \cdot K\left(\frac{u-u_l}{h(k)}\right)}{\sum_{l=1}^{k} K\left(\frac{u-u_l}{h(k)}\right)}$$

where K is so called *kernel function*.

Example – rectangular kernel

Rectangular kernel

Using rectangular (uniform) kernel, $K(u) = I_{[0,1)}(u)$, we obtain a simple estimate

$$\hat{\mu}(u) = \frac{\sum_{l \in L} y_l}{\#L} \text{ where } L = \{l : u_l \in [u - h(k), u + h(k))\}$$

- Other kernels: *Epanechnikov, Gauss, Cauchy...*

Convergence & rates

Convergence

If $h(k) \to 0$ and $k \cdot h(k) \to \infty$ then

$$\text{MISE}\,\hat{\mu} \to 0 \text{ as } k \to \infty.$$

Convergence rate

Let λ be a number of derivatives of μ. If $h(k) = k^{-\frac{1}{2\lambda+1}}$ then

$$\text{MISE}\,\hat{\mu} \sim k^{-\frac{2\lambda}{2\lambda+1}}.$$

- the smoother nonlinearity the faster convergence
- the rate can be established for smooth nonlinearities only

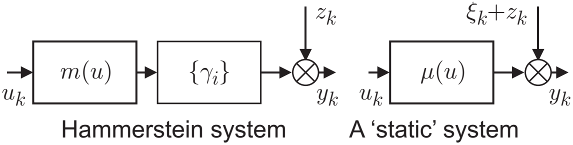
Class of systems

The class of systems to which the above algorithms can be **directly** includes many popular structures like:

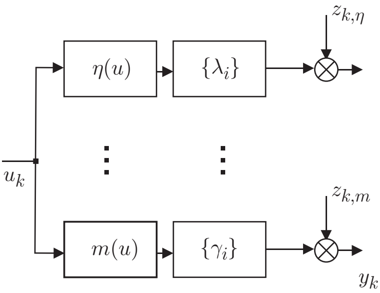
- *Hammerstein* system,
- parallel system,
- *Uryson* system, etc.

For instance, for *Hammerstein* system we have

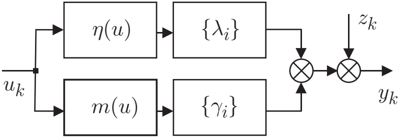
$$y_k = \overbrace{\gamma_0 m\left(u_k\right)}^{\mu(u_k)} + \overbrace{\sum_{i=1}^{\infty} \gamma_i \left[m\left(u_{k-i}\right) - Em\left(u_1\right)\right]}^{\xi_k} + z_k = \mu\left(u_k\right) + \xi_k + z_k$$



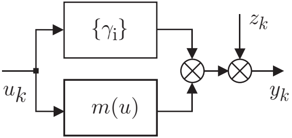
Examples of admissible dynamic nonlinear systems



Multichannel system



Uryson system



Parallel system

A censored (kernel) sample-mean approach to Wiener system identification

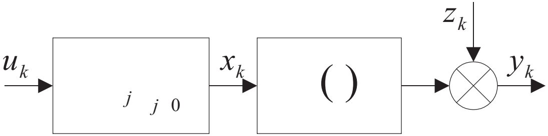


Figure: Wiener system

$$y_k = \mu \left(\sum_{j=0}^{\infty} \lambda_j u_{k-j} \right) + z_k$$

Assumptions

(A1) $\{u_k\}$ – i.i.d., bounded ($|u_k| < u_{\max}$) random process
(A1a) there exists a p.d.f. of the input $\vartheta_u(u_k)$, which is a continuous and strictly positive function in the estimation points x, i.e., $\vartheta_u(x) > 0$. or
(A1b) It holds that $P(u_k = x) > 0$ if u_k has discrete distribution.
(A2) The unknown impulse response $\{\lambda_j\}_{j=0}^{\infty}$ of the linear *IIR* filter is bounded from above as follows

$$|\lambda_j| \leqslant c \cdot \lambda^j$$

where $\lambda \in (0, 1)$ is a priori known constant.
(A3) The nonlinearity $\mu(x)$ is an arbitrary function, which is continuous almost everywhere on $x \in (-u_{\max}, u_{\max})$ (in the sense of Lebesgue measure).
(A4) The output noise $\{z_k\}$ is a zero-mean ergodic process, which is independent of the input $\{u_k\}$.

The algorithm

$$\widehat{\mu}_N(x) = \frac{\sum_{k=1}^{N} y_k \cdot K\left(\frac{\Delta_k(x)}{h(N)}\right)}{\sum_{k=1}^{N} K\left(\frac{\Delta_k(x)}{h(N)}\right)}$$

where

$$\begin{aligned} \Delta_k(x) \triangleq \; & |u_k - x|\,\lambda^0 + |u_{k-1} - x|\,\lambda^1 + |u_{k-2} - x|\,\lambda^2 + \ldots \\ & \ldots + \left|u_{k-S(N)-1} - x\right|\,\lambda^{S(N)-1} \end{aligned}$$

Parametric-nonparametric approach to Hammerstein system identification

Assumptions

A1: $|u_k| \leqslant u_{\max}$, $\exists$ p.d.f. $\nu(u)$

A2:

$$|\mu(u)| \leqslant w_{\max}$$

A3:

$$\sum_{i=0}^{\infty} |\gamma_i| < \infty$$

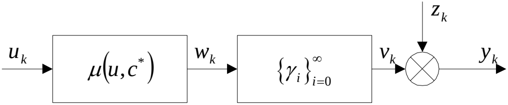
A4: $\mu(u_0)$ known for some u_0 (let $u_0 = 0$) and $\gamma_0 = 1$

Z5:

$$z_k = \sum_{i=0}^{\infty} \omega_i \varepsilon_{k-i}$$

$\{\varepsilon_k\}$ – i.i.d., independent of $\{u_k\}$, $E\varepsilon_k = 0$, $|\varepsilon_k| \leqslant \varepsilon_{\max}$
$\{\omega_i\}_{i=0}^{\infty}$ – unknown,
$\sum_{i=0}^{\infty} |\omega_i| < \infty$

Parameter knowledge



- we are given the formula $\mu(u,c)$, such that $\mu(u,c^*) = \mu(u)$, where $c^* = (c_1^*, c_2^*, ..., c_m^*)^T$ – true parameters
- $\mu(u,c)$ – differentiable with respect to c
- for each $u \in [-u_{\max}, u_{\max}]$

$$\|\nabla_c \mu(u,c)\| \leqslant G_{\max} < \infty, \qquad c \in \mathcal{C}(c^*)$$

- c^* is identifiable, i.e. there exist such sequence $\overline{u}_1, \overline{u}_2, ..., \overline{u}_{N_0}$ that

$$\mu(\overline{u}_n, c) = \mu(\overline{u}_n, c^*),\ n = 1,2,...,N_0 \Rightarrow c = c^*$$

Estimation of the static characteristic

$$Q_{N_0}(c) = \sum_{n=1}^{N_0} \left[w_n - \mu(\overline{u}_n, c)\right]^2 \qquad c^* = \arg\min_c Q_{N_0}(c)$$

Stage 1: On the basis on M pairs $\{(u_k, y_k)\}_{k=1}^{M}$, for N_0 fixed points $\{\overline{u}_n;\ n = 1, 2, ..., N_0\}$ estimate $\{w_n = \mu(\overline{u}_n, c^*);\ n = 1, 2, ..., N_0\}$

$$\widehat{w}_{n,M} = \widehat{R}_M(\overline{u}_n) - \widehat{R}_M(0)$$

Stage 2: Optimize the following criterion

$$\widehat{Q}_{N_0,M}(c) = \sum_{n=1}^{N_0} \left[\widehat{w}_{n,M} - \mu(\overline{u}_n, c)\right]^2$$

with respect to c and take the $\widehat{c}_{N_0,M}$ as the estimate of c^*.

Limit properties

If the system is identifiable then

$$\delta \cdot \|c - c^*\|^2 \leqslant Q_{N_0}(c) \leqslant D \cdot \|c - c^*\|^2$$

Theorem

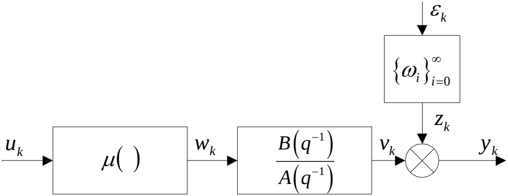
Assume that $\widehat{c}_{N_0,M}$ is unique and $\widehat{c}_{N_0,M}$, $c^* \in C$ for each M, where C is bounded convex set in R^m. If in stage 1

$$\widehat{R}_M(\overline{u}_n) = R(\overline{u}_n) + O(M^{-\tau}) \text{ in probability as } M \to \infty$$

for $n = 1, 2, ..., N_0$ and for $\overline{u}_n = 0$ then

$$\widehat{c}_{N_0,M} = c^* + O(M^{-\tau}) \text{ in probability as } M \to \infty$$

Prior knowledge of the linear dynamics



$$
\begin{aligned}
v_k &= b_0 w_k + ... + b_s w_{k-s} + a_1 v_{k-1} + ... + a_p v_{k-p} \\
\theta &= (b_0, b_1, ..., b_s, a_1, a_2, ..., a_p)^T \\
\vartheta_k &= (w_k, w_{k-1}, ..., w_{k-s}, y_{k-1}, y_{k-2}, ..., y_{k-p})^T \\
y_k &= \vartheta_k^T \theta + \bar{z}_k, \qquad \bar{z}_k = z_k - a_1 z_{k-1} - ... - a_p z_{k-p} \\
Y_N &= \Theta_N \theta + Z_N, \quad \Theta_N = (\vartheta_1, ..., \vartheta_N)^T, \quad Z_N = (\bar{z}_1, ..., \bar{z}_N)^T
\end{aligned}
$$

Nonparametric instrumental variables

$$\widehat{\theta}_{N,M}^{(IV)} = (\widehat{\Psi}_{N,M}^T \widehat{\Theta}_{N,M})^{-1} \widehat{\Psi}_{N,M}^T Y_N$$

where

$$\begin{aligned}
\widehat{\Theta}_{N,M} &= (\widehat{\vartheta}_{1,M}, \dots, \widehat{\vartheta}_{N,M})^T \\
\widehat{\vartheta}_{k,M} &= (\widehat{w}_{k,M}, \dots, \widehat{w}_{k-s,M}, y_{k-1}, \dots, y_{k-p})^T \\
\widehat{\Psi}_{N,M} &= (\widehat{\psi}_{1,M}, \dots, \widehat{\psi}_{N,M})^T \\
\widehat{\psi}_{k,M} &= (\widehat{w}_{k,M}, \dots, \widehat{w}_{k-s,M}, \widehat{w}_{k-s-1,M}, \dots, \widehat{w}_{k-s-p,M})^T
\end{aligned}$$

Limit properties (1)

Theorem

If the estimate $\widehat{R}_M(u)$ *is bounded, converges to* $R(u)$*, and the estimation error in the points* $u \in \{0, u_{k-r}$*; for* $k = 1, 2, ..., N$ *and* $r = 0, 1, ..., s+p\}$ *behaves like*

$$\left|\widehat{R}_M(u) - R(u)\right| = O(M^{-\tau}) \text{ in probability}$$

then for $NM^{-\tau} \longrightarrow 0$ *the following conditions are fulfilled*

(a') $Plim_{M,N\rightarrow\infty}\left(\frac{1}{N}\widehat{\Psi}^T_{N,M}\widehat{\Theta}_{N,M}\right)$ *exists and is not singular*

(b') $Plim_{M,N\rightarrow\infty}\left(\frac{1}{N}\widehat{\Psi}^T_{N,M}Z_N\right) = 0$

Limit properties (2)

Theorem

Under assumptions of Theorem 2 it holds that

$$\widehat{\theta}^{(IV)}_{N,M} \longrightarrow \theta \quad \textit{in probability}$$

as $N, M \longrightarrow \infty$, *if* $NM^{-\tau} \longrightarrow 0$. *In particular, for* $M \sim N^{(1+\alpha)/\tau}$, $\alpha > 0$, *the asympptotic rate of convergencehas the form*

$$\left\| \widehat{\theta}^{(IV)}_{N,M} - \theta \right\| = O(N^{-\min(\frac{1}{2},\alpha)}) \quad \textit{in probability}$$

Optimal instrumental variables

$$\Delta_N^{(IV)}(\Psi_N) \triangleq \widehat{\theta}_N^{IV} - \theta^*$$

$$Q(\Psi_N) \triangleq \max_{\left\|Z_N^*\right\|_2 \leq 1} \left\|\Delta_N^{(IV)}(\Psi_N)\right\|_2^2$$

$$Z_N^* \triangleq \frac{\frac{1}{\sqrt{N}} Z_N}{\bar{z}_{\max}}$$

Theorem

In Hammerstein system, for each admissible Ψ_N itholds that

$$\lim_{N\to\infty} Q(\Psi_N) \geqslant \lim_{N\to\infty} Q(\Psi_N^*) \text{ with probability } 1$$

where

$$\Psi_N^* = (\psi_1^*, \psi_2^*, ..., \psi_N^*)^T, \qquad \psi_k^* = (w_k, ..., w_{k-s}, v_{k-1}, ..., v_{k-p})^T.$$

Approximate realization

$$\widehat{\psi}^*_{k,M} = (\widehat{w}_{k,M}, \widehat{w}_{k-1,M}, \ldots, \widehat{w}_{k-s,M}, \widehat{v}_{k-1,M}, \widehat{v}_{k-2,M}, \ldots, \widehat{v}_{k-p,M})^T$$

$$\widehat{v}_{k,M} = \sum_{i=0}^{F} \widehat{\gamma}_{i,M} \widehat{w}_{k-i,M}$$

$$\widehat{\gamma}_{i,M} = \widehat{\varkappa}_{i,M} / \widehat{\varkappa}_{0,M}, \qquad \widehat{\varkappa}_{i,M} = \frac{1}{M} \sum_{k=1}^{M-i} (y_{k+i} - \overline{y})(u_k - \overline{u})$$

Summary

- Consistent estimates in the presence of colored noise
- Problem decomposition with use of nonparametric methods
- Broad class of models (non-linear-in-parameters + IIR)

Semiparametric algorithm – assumptions

- The nonlinear characteristic $m(u)$ can be an *arbitrary* function, square integrable, and *e.g.*:
  - differentiable
  - continuous
  - piecewise-smooth
- There is a set of input-output measurements $\{u_l, y_l\}$, $l = 1, \ldots, k$.
- There is a **polynomial** model $\mu_p(u)$ of order $p-1$ of the nonlinearity $\mu(u)$; *e.g.* hard-wired, or taken from *Matlab* System Identification toolbox.

Remark

The model can offer only crude approximations when the genuine nonlinearity turns out to be e.g. a piecewise smooth function with discontinuities.

Additive regression

- Having, by assumption, the polynomial model $\mu_p(u)$, we are interested in the remaining part:

$$\mu_r(u) = \mu(u) - \mu_p(u) = E(y_k | u_k = u) - \mu_p(u)$$

  which will further be referred to as *residual nonlinearity*.

- The polynomial model $\mu_p(u)$ can *exactly* be represented as a 'crude' wavelet approximation

$$\mu_p(u) = \sum_{i=0}^{p-1} \alpha_i \cdot u^i = \sum_{n=0}^{2^M-1} \alpha_{Mn}^p \cdot \varphi_{Mn}(u)$$

  where $\alpha_{Mn}^p = \langle \tilde{\mu}_p, \varphi_{Mn} \rangle$.

Wavelet estimate of a residual function

- The estimate is a version of the presented wavelet estimate

$$\hat{\mu}_r(u) = \sum_{n=0}^{2^M-1} \hat{\alpha}_{Mn} \cdot \varphi_{Mn}(u) + \sum_{m=M}^{K-1} \sum_{n=0}^{2^m-1} \hat{\beta}_{mn} \cdot \psi_{mn}(u)$$

  where the expansion coefficient estimates are computed in a convenient *on-line* fashion

$$\begin{bmatrix} \hat{\alpha}_{Mn}^{(k+1)} \\ \hat{\beta}_{mn}^{(k+1)} \end{bmatrix} = \begin{bmatrix} \hat{\alpha}_{Mn}^{(k)} \\ \hat{\beta}_{mn}^{(k)} \end{bmatrix} + (y_{k+1} - y_{l+1}) \begin{bmatrix} \Phi_{Mn}(u_{k+1}) - \Phi_{Mn}(u_l) \\ \Psi_{mn}(u_{k+1}) - \Psi_{mn}(u_l) \end{bmatrix}$$

  where $\Phi_{Mn}(u)$ and $\Psi_{mn}(u)$ are antiderivatives of $\varphi_{Mn}(u)$ and $\psi_{mn}(u)$.
- The algorithm starts with

$$\begin{bmatrix} \hat{\alpha}_{Mn}^{(1)} \\ \hat{\beta}_{mn}^{(1)} \end{bmatrix} = \begin{bmatrix} -\alpha_{Mn}^{p} \\ 0 \end{bmatrix} \text{ and } \{(u_0 = 0, y_0 = 0), (u_1 = 1, y_1 = 0)\}.$$

Convergence & rates

Convergence rate

Let λ be a number of derivatives of μ. If $K(k) = \frac{1}{2\lambda+1}\log_2 k$ then

$$\operatorname{MISE}\hat{\mu} \sim k^{-\frac{2\lambda}{2\lambda+1}}$$

Let $\mu(u)$ has a finite number of jumps. If $K(k) = \frac{1}{2}\log_2 k$ then

$$\operatorname{MISE}\hat{\mu} \sim k^{-\frac{1}{2}}$$

- the smoother nonlinearity the faster convergence (the **same** as for polynomials)
- the rate can be established also for **discontinuous** nonlinearities!
- the convergence holds regardless the actual type of the pre-model $\mu_p(u)$, be it regular or orthogonal.

Example - Legendre polynomial model and Haar wavelet amendment

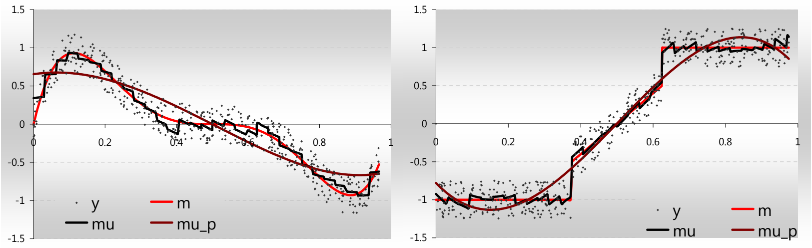
- The nonlinearities

$$m(u) = 5(u^5 - u^3) \text{ and } m(u) = \begin{cases} -1 & \text{if} \quad u < 3/8 \\ 4x - 2 & \text{if} \quad 3/8 \leq u < 5/8 \\ 1 & \text{if} \quad 5/8 \leq u \end{cases}$$

- The model (based on *Legendre* polynomial of order $p = 4$)

$$\mu_p(u) = \sum_{i=0}^{4} \alpha_i p_i(u) \text{ where } \alpha_i = \langle \mu_n, p_i \rangle$$

Example – simulation results



Final conclusions

- *Parametric* and *nonparametric* algorithms complete each other rather than compete…
- The choice of the algorithm type can separately be made appropriately to a different *a priori* knowledge available for either of the system block.
- The convergence of the algorithms can formally be shown for *virtually all* nonlinear characteristics.
- *Semiparametric* algorithms benefit from advantages of parametric and nonparametric ones.

A discovery of Ceres

Beginnings...

- Ceres was spotted by *G. Piazzi* as a result of an exhaustive search in an attempt to verify *Titius-Body* rule (**ad hoc model**) governing the distance of the Solar system objects from Sun).
- The observation of its position were recorded yet no orbit parameters had been established.
- The dwarf-planet was lost after traversing behind Sun.
- Several astronomers (*Body*, *von Zach*, *Olbers*) tried to determine the orbit and failed...

A discovery of Ceres

Towards better models. . .

- They used **a wrong model** (inappropriate *a priori* knowledge) assuming circular shape of the orbit (which result in *a biased model with systematic error*), and did also not correctly deal with error in measurements.
- Gauss ingeniously took into account these errors (proposing his *least squares* algorithm to cope with *random errors*) but also used a **better model** admitting elliptical orbits (*e.g.* the one based on *Kepler*'s laws).
- That the *Kepler*'s laws were not an **ultimate model** for celestial bodies motion was discovered and explained another 100 years later by another genius, *Albert Einstein*, whose *general relativity theory* finally explained Mercury's orbit anomalies.

Selected recent papers of the team

W. Greblicki.
Continuous-time Hammerstein system identification.
IEEE Transactions on Automatic Control, 2000.

Z. Hasiewicz.
Non-parametric estimation of non-linearity in a cascade time series system by multiscale approximation.
Signal Processing, 2001.

W. Greblicki.
Nonlinearity recovering in Wiener system driven with correlated signal.
IEEE Transactions on Automatic Control, 2004.

Z. Hasiewicz and **G. Mzyk**.
Combined parametric-nonparametric identification of Hammerstein systems.
IEEE Transactions on AC, 2004.

Z. Hasiewicz and **G. Mzyk**.
Hammerstein system identification by nonparametric instrumental variables.
International Journal of Control, 2008.

Z. Hasiewicz, M. Pawlak, and **P. Śliwiński**.
Non-parametric identification of non-linearities in block-oriented complex systems by orthogonal wavelets with compact support.
IEEE Transactions on CAS I, 2005.

G. Mzyk.
A censored sample mean approach to nonparametric identification of nonlinearities in Wiener systems.
IEEE Transactions on CAS II, 2007.

M. Pawlak, Z. Hasiewicz, and P. Wachel.
On nonparametric identification of Wiener systems.
IEEE Transactions on SP, 2007.

P. Śliwiński and Z. Hasiewicz.
Computational algorithms for multiscale identification of nonlinearities in Hammerstein systems with random inputs.
IEEE Transactions on SP, 2005.

P. Śliwiński and Z. Hasiewicz.
Computational algorithms for wavelet identification of nonlinearities in Hammerstein systems with random inputs.
IEEE Transactions on SP, 2008.